6.867 — Lecture 1 Poll Deck: Classification & Perceptron

Ayush Khaitan

Q1 — What counts as "data" in this setup?

Context. We build an access system from face images labeled $+1$ (permit) or -1 (deny).

Question. What is the *minimal* resource needed to learn a classifier here?

A) **A**. Unlabeled images only

B) **B**. Labeled pairs (x, y)

C) **C**. Only positive $(+1)$ examples

D) **D**. Only negative (-1) examples

q1

Q2 — Dimensionality of image vectors

Context. A grayscale image is turned into a column vector $x \in \mathbb{R}^d$ by stacking pixels.

Question. For 100×100 images, what is d?

A) **A**. 100

B) **B**. 200

C) **C**. 10,000

D) **D**. 1,000,000

Q3 — The real objective

Context. A rule can memorize the training images yet fail on new ones.

Question. What is the actual goal emphasized in Lecture 1?

A) **A**. Zero training error

B) **B**. Minimize runtime only

C) **C**. Strong generalization to unseen data

D) **D**. Memorize with hashes

q3

Q4 — Model class too large

Question. If the hypothesis class is too rich, what is the typical risk?

A) **A**. Underfitting

B) **B**. Overfitting (poor generalization)

C) **C**. Only slow optimization

D) **D**. Automatic data leakage

Q5 — Model class too small

Question. If the hypothesis class is too restrictive, the likely outcome is:

A) **A**. Underfitting

B) **B**. Data poisoning

C) **C**. Perfect training accuracy

D) **D**. Guaranteed robustness

Q6 — Which learning paradigm?

Context. Each image has a label in $\{\pm 1\}$.

Question. This setup is:

A) **A**. Unsupervised learning

B) **B**. Supervised *classification*

C) **C**. Regression

D) **D**. Reinforcement learning

Q7 — Linear decision rule

Question. A linear classifier “through the origin” predicts:

A) **A**. $\mathrm{sign}(b + \theta^\top x)$ with $b \neq 0$

B) **B**. $\mathrm{sign}(\theta^\top x)$

C) **C**. $\theta^\top x$ without sign

D) **D**. Random label

q7

Q8 — When does perceptron update?

Question. The perceptron weight update fires when:

A) **A**. The prediction is correct

B) **B**. A mistake occurs

C) **C**. A timer elapses

D) **D**. Epoch ends

q8

Q9 — Update formula

Question. The standard perceptron update on a mistake is:

A) **A**. $\theta \leftarrow \theta - yx$

B) **B**. $\theta \leftarrow \theta + yx$

C) **C**. $\theta \leftarrow yx$

D) **D**. $\theta \leftarrow \theta + x$

q9

Q10 — Role of the label y

Question. In $\theta \leftarrow \theta + yx$, the label y primarily:

A) **A**. Determines the *sign* of the update

B) **B**. Scales magnitude independently

C) **C**. Does nothing

D) **D**. Randomizes the step

Q11 — High-dimensionality

Question. As d grows large, a basic risk for simple linear models is:

A) **A**. Guaranteed accuracy

B) **B**. Overfitting if capacity is unchecked

C) **C**. Parameter scarcity

D) **D**. No need for labels

Q12 — Useful negatives

Context. Few “deny” examples in the collected data.

Question. Adding images of other people captured with similar camera orientation helps because:

A) **A**. They are identical to positives

B) **B**. They enrich the negative distribution without shifting features

C) **C**. They add labeling noise

D) **D**. They force underfitting

q12

Q13 — Training error

Question. The 0–1 training error used in Lecture 1 counts:

A) **A**. Squared deviations

B) **B**. Margin violations only

C) **C**. Misclassifications ($y \neq f(x)$)

D) **D**. Only false negatives

q13

Q14 — A limitation of linear pixels

Question. Which statement is true for a linear classifier on raw pixels?

A) **A**. A fixed permutation of pixel order (applied to all data) leaves predictions unchanged

B) **B**. It models local edges explicitly

C) **C**. It captures 2D geometry intrinsically

D) **D**. It always needs convolution

Q15 — Geometric view

Question. The boundary $\theta^\top x = 0$ is:

A) **A**. A hyperplane with normal vector θ

B) **B**. A sphere centered at θ

C) **C**. A paraboloid

D) **D**. A decision tree

Q16 — Key perceptron identity

Question. After an error update $\theta' = \theta + yx$, which identity holds?

A) **A**. $y\,\theta'^{\top}x = y\,\theta^{\top}x + \|x\|^2$

B) **B**. $\theta'^{\top}x = \theta^{\top}x$

C) **C**. $y\,\theta'^{\top}x = -\,y\,\theta^{\top}x$

D) **D**. $\|\theta'\| = \|\theta\|$

Q17 — Class imbalance

Question. With very few negatives, a basic mitigation is to:

A) **A**. Ignore negatives entirely

B) **B**. Relabel positives as negatives

C) **C**. Collect more representative negatives

D) **D**. Discard positives

Q18 — What is generalization error?

Question. Generalization error is measured on:

A) **A**. The training set

B) **B**. Fresh samples drawn from the task distribution

C) **C**. Random noise inputs

D) **D**. Only the hardest cases

Q19 — Why not a “pixel-i” rule?

Context. One can contrive a rule using a single pixel index to fit the training set exactly.

Question. Why is this poor practice?

A) **A**. It’s too slow

B) **B**. It fails to generalize

C) **C**. It breaks linear algebra

D) **D**. It requires labels

Q20 — Interpreting θ

Question. In a linear classifier through the origin, θ is best viewed as:

A) **A**. The hyperplane's normal

B) **B**. The data mean

C) **C**. A covariance matrix

D) **D**. A bias term

Q21 — Pixel permutation invariance

Question. If we apply the *same* fixed pixel permutation to every image at train and test time, a linear classifier's predictions:

A) **A**. Are unchanged up to permuting the weights

B) **B**. Become random

C) **C**. Exploit spatial locality

D) **D**. Always improve

Q22 — Loss function character

Question. Which loss used for training error is discrete and non-differentiable?

A) **A**. Zero–one loss

B) **B**. Squared loss

C) **C**. Logistic loss

D) **D**. Hinge loss

Q23 — When does perceptron converge?

Question. Classic perceptron convergence (finite mistakes) is guaranteed when:

A) **A**. Data are nonlinearly separable

B) **B**. Data are linearly separable (positive margin)

C) **C**. Labels are random

D) **D**. $\theta = 0$ forever

Q24 — "Through the origin" means...

Question. Saying the classifier is "through the origin" implies:

A) **A**. The boundary is $\theta^\top x = b$ with $b \neq 0$

B) **B**. No bias term; boundary is $\theta^\top x = 0$

C) **C**. $\theta = 0$

D) **D**. Only spherical data work

Q25 — A quick sanity check

Question. For 28×28 grayscale images, the vector dimension is:

A) **A**. 56

B) **B**. 112

C) **C**. 784

D) **D**. 28